Joint Optimization of Caching Placement and Power Allocation in Virtualized Satellite-Terrestrial Network

Haijun Zhang, *Fellow, IEEE*, Jiali Xu, Xiangnan Liu, Keping Long, *Senior Member, IEEE*, and Victor C. M. Leung, *Life Fellow, IEEE*

***Abstract*— With the rapid development of mobile services and applications, the transmitting of massive data makes low-cost communication a challenge. Edge-based wireless communication technology is developed to be a promising approach to satisfy the communication requirements. Edge caching technology is one of effective methods to reduce the overhead of communication system and the pressure of backhauls. In this paper, the joint optimization problem of caching placement and power allocation in virtualized low earth orbit (LEO) satellite-terrestrial networks is proposed, which is based on cooperative caching, by considering cache size limits and power constraints. The optimization problem is solved using an algorithm inspired by the courtship movements and random flights of mayflies. Simulation results show the effectiveness of the proposed scheme in improving system performance and reducing power consumption.**

***Index Terms*— Virtual resource allocation, satellite-terrestrial networks, caching placement, power allocation.**

I. INTRODUCTION

WITH the rapid development of the Internet of Things (IoT) and 5G mobile communication technologies, various mobile services and applications such as online games, webcasting, and virtual reality have become extremely common [1], [2], [3]. The number of mobile devices has boomed in recent decades, which makes it a challenge transmitting the massive mobile data effectively.

In order to handle huge traffic requests and relieve the cellular networks from the pressure on the backhaul, technologies such as mobile edge computing (MEC) and device to device (D2D) communication have become research hotspots [4], [5], [6], [7], [8], [9], [10]. By deploying MEC servers on edge nodes (such as base stations, relays, content routers, etc.) in networks, massive amounts of data no longer needs to be uploaded to the cloud for processing, and can be processed at the edge of the network. Mobile edge caching can be realized through the storage function of the servers, which provides support for decreasing the data traffic of the core network, relieving backhaul pressure, and reducing the latency so as to alleviate the shortage of communication resources [11], [12], [13]. The edge cache system with D2D can make use of the storage space of user devices [14], [15]. Users obtain the required content directly from the surrounding users through D2D links without going through the cellular backhaul links. In this way, the load of the base station is reduced and the energy efficiency and scalability of networks are improved.

Currently, mobile communication research has come to the new development stage of B5G. Though the cover area of communication networks has been expanded and the throughput of the entire system has been improved with the development of recent heterogeneous networks [16], there are still some intractable problems. Due to the scarcity of communication resources and the difficulty of deploying access points (APs), especially for remote areas, communication networks coverage is affected [17]. The limited backhaul capacity of small cells may cause degradation of the overall communication system performance, which is negative for reducing the cost of communication system. With the advancement of satellite technologies in an all-round manner, including satellite launch, satellite manufacturing, integrated circuits, and communication technology [18], [19], the satellite-terrestrial networks that have been widely discussed are expected to be applied in many fields [20]. The integration of ground mobile communication and satellite communication can better meet terminals' communication requirements and weaken the restrictions of geographical conditions, which is a necessary supplement and extension of cellular mobile communication. From the perspective of cost, micro-satellite platforms are usually adopted in LEO satellites communication. The technical difficulty and satellite scale of LEO satellites communication are lower than those of satellites communication system with high orbits, and the cost of single-satellite development is significantly reduced. For traditional terrestrial communication network

Manuscript received 31 March 2022; revised 17 September 2022 and 11 January 2023; accepted 6 March 2023. Date of publication 20 March 2023; date of current version 13 November 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFB1806103, in part by the National Natural Science Foundation of China under Grant 62225103 and Grant U22B2003, in part by the Beijing Natural Science Foundation under Grant L212004, and in part by the China University Industry-University-Research Collaborative Innovation Fund under Grant 2021FNA05001. The associate editor coordinating the review of this article and approving it for publication was Z. Fadlullah. *(Corresponding author: Haijun Zhang.)*

Haijun Zhang, Jiali Xu, Xiangnan Liu, and Keping Long are with the Beijing Engineering and Technology Research Center for Convergence Networks and Ubiquitous Services, University of Science and Technology Beijing, Beijing 100083, China (e-mail: haijunzhang@ieee.org; xujiali@xs.ustb.edu.cn; xiangnan.liu@xs.ustb.edu.cn; longkeping@ustb.edu.cn).

Victor C. M. Leung is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China, and also with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: vleung@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2023.3257178.

Digital Object Identifier 10.1109/TWC.2023.3257178

developers, the difficulty of popularizing the communication networks in remote areas lies in equipment cost, data cost, service cost and so on. It is promising to find the cost control solutions for the construction of space segment and ground station, which means that the application prospect is broad [17].

The satellite-terrestrial networks can provide terminals with more flexible communication services, and the related research based on LEO satellite-terrestrial communication is also gradually advancing [21], [22], [23], [24], [25], [26], [27], [28], [29]. Some researchers proposed that LEO satellite could act as a destination relay between other LEO satellites and users [22]. Some researchers proposed a satellite-terrestrial based networking architecture, which combined LEO satellites networks and the terrestrial networks for the transmission of various communication services [24]. Some researchers proposed a scheme that each user uploaded its data through a macro station, a traditional small station or a LEO-based small station. The terrestrial stations uploaded the traffic to the core network through the matching links [25].

There is no doubt that satellites play a promising role in the design of high-performance networks. However, the LEO satellite-terrestrial networks also face many challenges [26]. For example, the transmission type of satellites based on laser or microwave is easily affected by the environment. Compared with the terrestrial networks with large bandwidth and mature routing mechanism, the routing mechanism of LEO satellite network is not stable enough. Another important point is the increasing QoS requirements, such as high data rates and low processing power consumption, which may limit the application of satellite-terrestrial networks. The application of MEC that has been deeply researched in the terrestrial cellular network in the satellite-terrestrial networks is undoubtedly an effective solution to effectively meet the communication requirements of terminals.

In the satellite-terrestrial network architecture that integrates MEC, network function virtualization (NFV) is a potential technical solution to improve network flexibility. With the development of B5G communication technologies and the extension of communication services, mobile virtual network operators (MVNOs) that closely meet the demands of customers and focus on data services have gained widespread attention in the market [30], [31], [32]. NFV is an operational framework designed to orchestrate virtualized network function (VNF) software devices and perform full lifecycle management of VNF devices. It has a positive effect on breaking the barriers of special network with dedicated hardware and greatly reducing the cost of network construction for operators. By using NFV technology to unify available resources, the satellite-terrestrial networks integrating MEC can centrally manage offloading tasks. NFV has huge advantages in reducing the cost and time of deploying services for network providers in satellite-terrestrial networks with MEC.

From the perspective of MVNOs, it is meaningful to consider how to reduce the operation cost of communication networks with effective communication technologies, which means low-cost communication systems are required. On the one hand, the pressure caused by huge data traffic in the core network should be shunted away efficiently. On the other hand, in satellite-terrestrial networks, reducing power consumption is necessary for the whole system because of the limited energy storage and insufficient energy supply compared with terrestrial communication systems [33], [34]. In view of this situation, proceeding with further research on efficient wireless resource allocation strategy is necessary for the communication system. Although some researches on resource allocation in satellite-terrestrial networks have been carried out [35], [36], [37], [38], [39], [40], [41], [42], there are few solutions that consider resource allocation including cache resource and power resource jointly. Besides, with the diversification of network functions and the complication of communication equipment, virtual network resource allocation in satellite-terrestrial networks also needs to be studied further. With the application of NFV technology, various infrastructures, communication resources, computing resources, and storage resources in the network can be abstracted to build satellite-terrestrial integration resources. These virtualized resources are pooled together to provide global information about network resources. Then how to optimize the resource allocation of the system is a challenging subject.

Many intelligent optimization algorithms have been invented to solve such complex optimization problems [43], [44], [45], [46], [47], [48], [49]. The study in [43] designed a collaborative content placement algorithm based on ant colony optimization, which could achieve fine-grained collaborative content layout close to the optimal performance. The researchers in [44] proposed a combination of genetic algorithm and particle swarm algorithm for the joint optimization of load decision, spectrum, and power allocation. The works in [45] designed a three-layer optimization framework based on particle swarm algorithm to reduce MEC delay and power consumption. The resource allocation solution based on particle swarm optimization algorithm for the dynamically changing fog computing platform was proposed in [46]. The study proposed a resource management algorithm based on modified particle swarm optimization algorithm to assign tasks in an effective way [47]. The researchers designed a new resource allocation strategy based on the ant colony algorithm framework for the downlink orthogonal frequency division multiple access network [48]. These studies have demonstrated the effectiveness of swarm intelligent algorithms in resource optimization, but most of them are based on classical algorithms such as ant colony algorithm and particle swarm optimization. The development of new algorithms also provides more solutions for resource optimization in communication networks.

Although the researches on satellite-terrestrial networks are thriving, there are few studies on the virtualization framework under the satellite-terrestrial network. This paper discusses the satellite-terrestrial virtualization framework with MEC, where the control layer of the network can work logically in a centralized manner and physically in a distributed manner. In the previous researches, most of the optimization of resources such as caching and communication were considered separately.

In this paper, a coding-based cooperative way to provide cached content is proposed, and the joint optimization problem of virtualized resources is considered in the satellite-terrestrial network with MEC caching. In addition, this paper adopts a newly proposed mayfly algorithm [50] to solve the joint resource optimization problem. The main contributions of this paper can be summarized as follows.

- Design a coding-based collaborative cache scheme based on MEC for the transmission of data in LEO satellite-terrestrial networks. The resources such as cache in the satellite-terrestrial network architecture are made closer to terminals. Based on the advantages of LEO satellite relay transmission, terminals in remote areas can also be served. Terminals obtain contents from local servers, D2D links, and cooperative terrestrial APs, while LEO satellites can supplement the required content when the terminal is far from surrounding APs.
- Propose a virtualization framework in the satellite-terrestrial network. It enables the network that integrates MEC to logically and centrally manage offloading tasks. The income and expenditure of the entire communication system are considered to manage the cache and power resources to improve the performance of the communication network. The optimization process is mainly operated in the terrestrial center.
- Set different income and expenditure coefficients to form the system utility function. Unlike most studies that consider resources such as caching and communication separately, the joint optimization of the data traffic and the system power consumption is considered, with the constraints of the server memory and transmit power. Then the penalty function method is applied to convert the objective function with constraints. The mayfly algorithm is utilized to solve the optimization problem to obtain the resource allocation strategy.
- Carry out simulations with different parameter settings. The simulation results demonstrate that the proposed scheme works well in the overall system performance.

The rest of this paper is organized as follows. In Section II, the virtualization framework is investigated in wireless satellite-terrestrial networks and the system model with caching and power consumption is considered. In Section III, joint problem of caching and power allocation is formulated. In Section IV, we discuss the solution of the optimization problem based on the mayfly algorithm. Section V presents the simulation results and discussions to demonstrate the superiority of the proposed scheme. Finally, we conclude this paper in Section VI.

II. System Model

A. Satellite-Terrestrial Virtualization Framework

In the satellite-terrestrial network with MEC cache function, Fig. 1 enumerates the resource allocation parts that this paper will focus on in the following research. The physical resources include cache resources and communication resources, and these resources can be abstracted into virtual resource.

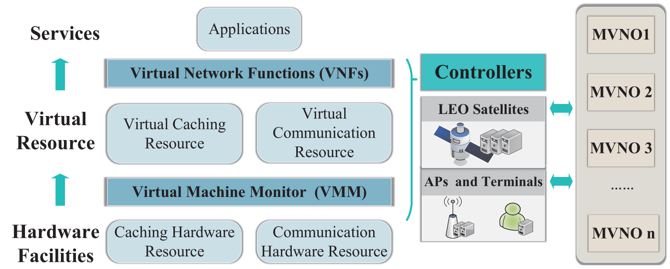


Fig. 1. The proposed satellite-terrestrial virtualization framework.

The virtual machine monitor (VMM) undertakes the responsibilities for managing physical resources, and is also responsible for the creation and management of the virtual environment. Various network services are supported by VNFs, ensuring that the virtualized network framework is flexible enough to meet different requirements. The logical separation and combination of resources can break the physical resource isolation, which is beneficial to maximize the utilization of physical hardware resources. In this way, operators can customize services according to terminals' requirements, and terminals can obtain better services.

According to the definition of European Telecommunications Standards Institute (ETSI), the NFV framework is composed of network service and management and orchestration [51]. In the view of network service, NFV infrastructure layer mainly includes the hardware and software resources that constitute the NFV environment, and virtual network function layer implements network functions by loading software on the virtual machines. The management and orchestration module is responsible for coordinating and monitoring the resources and functional operations of the entire platform. In the proposed framework, the resource allocation strategies are formulated by orchestrators.

In this virtual framework, the functions of control layer are abstracted as satellite and ground controllers. In this way, a control layer that works logically in a centralized manner and physically in a distributed manner is formed. These controllers are responsible for controlling and managing resource pools and assigning resources to terminals on demand. They receive terminal requirements, exchange information with network orchestrators, and obtain resource allocation strategies to allocate resources. Terminals in need receive virtualization services directly from virtualization operators. In the satellite-terrestrial network, the comprehensive processing of resource allocation strategies is implemented on the ground servers with powerful computing ability.

B. Network Model

In this paper, the situation that a single satellite provides services to the ground area is considered, which means terminals can only require services from a single satellite covering the area. In reality, due to the large number of satellites in densely deployed LEO networks [24], the same location can be covered by multiple satellites. LEO satellites are assumed to operate in the specified orbit, which means that each LEO satellite under unified control can periodically provide services

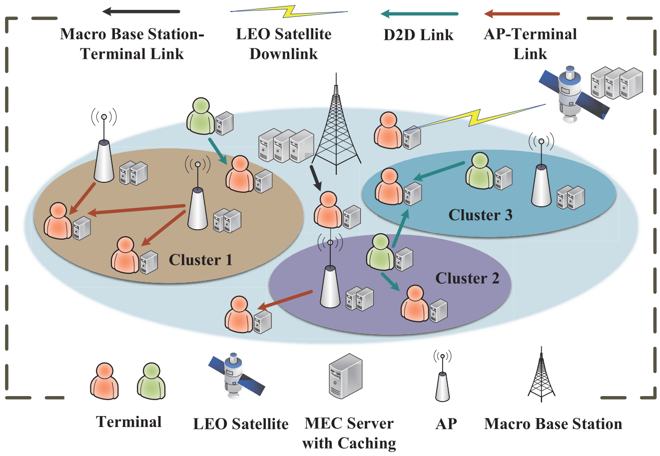


Fig. 2. The satellite-terrestrial network system model.

to the same ground area. Therefore, the case of communicating with single LEO satellite is considered [52]. As shown in Fig. 2, there are densely deployed APs, a macro base station, and terminals in the satellite coverage area.

Only the macro base station exchanges data with the core network through reliable optical fiber backhaul links. It is assumed that the communication of the ground system uses the C-band, and Ka-band is used when communicating with the LEO satellite. When distributing content, terminals in need can obtain cached contents from different devices. The specific cache acquisition method is described in the subsection of content request and cooperative delivery model. In this paper, the optimization problem of cache traffic and power is considered. Set $K = \{1, 2, \ldots, \mathrm{K}\}$ to represent the clusters, $J = \{1, 2, \ldots, \mathrm{J}\}$ to represent the APs and $I = \{1, 2, \ldots, \mathrm{I}\}$ to represent terminals. Considering the complexity of channel estimation with a large number of antennas, the perfect channel state information condition is assumed [53].

Denote $d_{ii'}$ as the distance between terminal i and i'. The maximum distance constraint of D2D links is denoted by Dis^d. Then the constraint is given by $d_{ii'} < Dis^d$. The signal to interference and noise ratio (SINR) from terminal i' to terminal i without considering the interference from APs is denoted by

$$\gamma_{ii'} = \frac{g_{ii'} p_d}{\sigma_d^2 + \sum\limits_{i'' \in I, i \neq i'', i' \neq i''} g_{ii''} p_d}, \tag{1}$$

where the p_d is the transmit power of terminals. $g_{ii'}$ and $g_{ii''}$ are the channel gains of the D2D and interfering links, respectively. The Additive White Gaussian Noise (AWGN) is denoted by $N_d \sim CN\left(0, {\sigma_d}^2\right)$ and ${\sigma_d}^2$ is the noise variance. The SINR from the AP j to terminal i is denoted by [54]

$$\gamma_{ij} = \frac{g_{ij} y_{ij} P_k^b}{\sigma_b^2 + \sum\limits_{j' \in J, j' \neq j} g_{ij'} y_{ij'} P_k^b}, \tag{2}$$

where $y_{ij} \in [0, 1]$ is the power allocation coefficient allocated to terminal i by AP j. P_k^b is the transmit power of APs in cluster k. g_{ij} and $g_{ij'}$ are the channel gains of the AP-terminals and interfering links, respectively. $y_{ij'}$ is the power allocation coefficient allocated to terminal i by AP j'. We assume that the transmit power of all APs in one cluster is the same. ${\sigma_b}^2$ is the noise variance. The constraint of power allocation coefficient is shown as

$$y_{ij} \in [0, 1], \quad \forall k \in K,\ \forall i \in I, \forall j \in J. \tag{3}$$

C. Coded Caching Placement Model

The coding used in storage has become the key problem of the system and many researchers have carried out researches on the storage coding. The files can be divided into k original blocks, and the blocks are encoded and then stored in multiple nodes. When the original data needs to be restored, k code blocks can be obtained from these nodes, and the original data can be restored by using the k data blocks for decoding operations. In practice, it can be achieved with Maximum-Distance Separable (MDS) codes [55], [56], [57]. In this paper, denote $F = \{f_1, f_2, \ldots, f_N\}$ as a set with N files with the same size. Each file in the set is divided into a equal-sized original blocks. The original blocks are encoded into encoded blocks to be cached, and the original file can be recovered at the received terminal when the number of received blocks re_{num} is more than or equal to the number of blocks of the original file ($re_{num} \geq a$).

Based on the symmetry premise of the MDS codes, we consider a series of transmission clusters deployed in the coverage area of the LEO satellite with a macro base station. APs and terminals are deployed in every cluster. In the same cluster, the content blocks cached in APs are the same due to the symmetry. It is worth noting that the local caching of terminals in the same cluster is also the same due to the symmetry. In addition, the cache list of the file is known globally.

Let L, L^b, and L^s denote the cache size of local server, AP server, and LEO satellite server, respectively. The storage performance of APs and LEO satellites is better than that of local servers actually. For the file n, let c_{kn}^d, c_{kn}^b, and c_n^s denote the cached blocks of terminals in cluster k, the cached blocks of APs in cluster k, and the cached blocks in the LEO satellite, respectively. The radio access network time slot is t. Then the transmit power of LEO satellites is given as [52]

$$P_n^s = {\sigma_s}^2 \left(\frac{4\pi d^s}{\lambda G}\right)\left(2^{\frac{c_n^s Z}{W^s t}} - 1\right), \tag{4}$$

where G is the antenna gain, λ is the wavelength, d^s is the distance between the satellite and the terminal i, W^s is the corresponding bandwidth, and ${\sigma_s}^2$ is the noise variance. The numbers of cache blocks cached by local servers, APs, and the LEO satellite are no more than that of corresponding device cache capacities, which is given as

$$\sum_{n \in N} c_{kn}^d \leq L, \quad \forall k \in K, \tag{5}$$

$$\sum_{n \in N} c_{kn}^b \leq L^b, \quad \forall k \in K, \tag{6}$$

$$\sum_{n \in N} c_n^s \leq L^s. \tag{7}$$

D. Content Request and Cooperative Delivery Model

There are two parts to be designed in the caching model: the content request and the content delivery.

For the request probability of contents, the content popularity is defined by Zipf distribution referring to related caching work [58]. θ_n is used to represent the request probability of terminals for the file n, and μ is an exponential feature describing the distribution. Specifically, $\mu = 0$ means a uniform distribution of content popularity. The request probability is defined as

$$\theta_n = \frac{1/n^{\mu}}{\sum_{n=1}^{N} 1/n^{\mu}}. \tag{8}$$

The sum of the probabilities of requesting file in each cluster is 1.

$$\sum_{n \in N} \theta_n^k = 1, \quad \forall k \in K. \tag{9}$$

For the content delivery period, a cooperation delivery scheme that allows terminals to get content from MEC servers is proposed. Each MEC server caches coded blocks without repeating, which is given as

$$c_{kn}^d \le a, \quad \forall k \in K,\ \forall n \in N, \tag{10}$$

$$c_{kn}^b \le a, \quad \forall k \in K,\ \forall n \in N, \tag{11}$$

$$c_n^s \le a, \quad \forall n \in N. \tag{12}$$

The binary cache decision variable $x \in \{0,1\}$ is set to represent the content acquisition case. When the terminal needs to acquire content, the different cases are listed as follows.

- Case 1: From local caching. $c_{kn}^d = a$ means the terminal can get required content from the local server directly without communicating with other terminals. $x_{in}^l = 1$ means terminal i gets the cached blocks of file n in local.
- Case 2: From D2D links. When $c_{kn}^d < a$, the terminal needs to receive blocks from other terminals. If nearby terminals with $\gamma > threshold$ can provide enough blocks more than a, D2D links can be established. $x_{in}^d = 1$ means terminal i gets the cached blocks of file n over D2D links.
- Case 3: From AP-terminal links. If the first two ways fail to provide enough blocks to recover the objective file, the terminal can obtain the content from APs that cache the related blocks and meet the decoding threshold conditions. Connectable APs can transmit the file data to the terminal at the same time. $x_{in}^b = 1$ means terminal i gets the cached blocks of file n from APs.
- Case 4: From satellite-terrestrial links. Remote terminals with adverse channel condition on the ground can also obtain data from the LEO satellite with cached blocks. $x_{in}^s = 1$ means terminal i gets the cached blocks of file n from the LEO satellite.

Over all, the case will be selected on the basis of the different content requests. APs and the LEO satellite send data to the terminal in the downlink while using different spectrums so as to avoid interference. Besides, the established D2D links between terminals transmit data through the uplink. And

$$x_{in}^l + x_{in}^d + x_{in}^b + x_{in}^s = 1, \quad \forall i. \tag{13}$$

III. Problem Formulation

When $c_{kn}^d \ge a$ in cluster k, the traffic that terminal i obtains from local device is given as

$$r_{in}^l = x_{in}^l c_{kn}^d. \tag{14}$$

To reduce the complexity and unnecessary overhead, we consider a distance constraint $d_{ii'} < Dis^d$ that terminals can only connect to others within a certain range to obtain cached content over D2D links, and the decoding threshold is ignored. Denote I_b^i as a set of terminals that provide D2D transmission for terminal i. $\sum_{i \in I_b^i} c_{kn}^d \ge a$ means that terminal i can obtain enough cache blocks of file n from D2D links and restore the file successfully. The offloading traffic of file n that terminal i obtains from D2D links is given as

$$r_{in}^d = \sum_{i \in I_b^i} x_{in}^d c_{kn}^d. \tag{15}$$

When the first two ways fail, terminals can connect to APs or the LEO satellite to obtain the blocks based on the broadcast caching list. Data can be provided by nearby APs with the constraints of distance. The decoding threshold is ignored. Denote d_{ij} as the distance between terminal i and AP j. The distance constraint between them is denoted by $d_{ij} < Dis^b$. J_b^i is the collection of eligible APs. $\sum_{j' \in J_b^i} c_{kn}^b \ge a$ means the APs can provide enough cached blocks to terminal i to obtain file n, and the offloading traffic of file n that terminal i obtains from APs is given as

$$r_{in}^b = \sum_{j \in J_b^i} x_{in}^b c_{kn}^b. \tag{16}$$

Apart from choosing nearby APs with objective coded blocks, the terminals that are not satisfied or far away from APs with adverse channel condition can connect to the satellite. The LEO satellite with related caching can provide the terminal with coded blocks of the required content while obtaining the rest of blocks from the core network through the backhaul link and sending them to terminals. For file n, the traffic transmitted through the backhaul link is $x_{in}^s (a - c_n^s)$. The offloading traffic of file n that terminal i obtains from the LEO satellite is given as

$$r_{in}^s = x_{in}^s c_n^s. \tag{17}$$

In the proposed model, the power consumption is composed of content caching and cooperation content transmitting. In particular, backhaul content transmitting consumption is not considered because of the minor impact. The power consumption of caching is given as

$$P_{ca} = P_{ca}^d + P_{ca}^b + P_{ca}^s$$

$$= Zp_{ca}\sum_{n\in N}\sum_{k\in K}\sum_{i\in I}\frac{c_{kn}^{d}}{a} + Zp_{ca}\sum_{n\in N}\sum_{k\in K}\sum_{j\in J}\frac{c_{kn}^{b}}{a} + Zp_{ca}\sum_{n\in N}\frac{c_{n}^{s}}{a}, \tag{18}$$

where p_{ca} is the power of caching per bit and is calculated in Watt. The size of every block is denoted by Z bits. P_{ca}^{d} is the power consumption of local caching. P_{ca}^{b} is the power consumption of caching in AP servers. P_{ca}^{s} is the power consumption of caching for the LEO satellite server.

The power consumption of transmitting is given as

$$\begin{aligned} P_{trans} &= P_{trans}^{d} + P_{trans}^{b} + P_{trans}^{s} \\ &= \sum_{n\in N}\sum_{k\in K}\theta_n^k\sum_{i\in I}x_{in}^{d}\sum_{i'\in I_b^i}p_d \\ &\quad + \sum_{n\in N}\sum_{k\in K}\theta_n^k\sum_{i\in I}\sum_{j'\in J_b^i}x_{in}^{b}y_{ij'}P_k^b \\ &\quad + \sum_{n\in N}\sum_{k\in K}\theta_n^k\sum_{i\in I}x_{in}^{s}P_n^s, \end{aligned} \tag{19}$$

where P_{trans}^{d} is the power consumption of transmitting by D2D links. P_{trans}^{b} is the power consumption of transmitting by APs. P_{trans}^{s} is the power consumption of transmitting by the LEO satellite. The maximum transmit power constraint for APs is shown as

$$\sum_{k\in K}\sum_{i\in I}y_{ij}P_k^b \le P_{\max}^{b}, \quad \forall j\in J. \tag{20}$$

where $P_{\max}^{b}$ is the maximum transmit power of APs. The total system traffic revenue that terminals obtain by MEC caching is computed as

$$\begin{aligned} Tra &= T^l + T^d + T^b + T^s \\ &= \sum_{n\in N}\sum_{k\in K}\sum_{i\in I}\theta_n^k r_{in}^{l} + \sum_{n\in N}\sum_{k\in K}\sum_{i\in I}\theta_n^k r_{in}^{d} \\ &\quad + \sum_{n\in N}\sum_{k\in K}\sum_{i\in I}\theta_n^k r_{in}^{b} + \sum_{n\in N}\sum_{k\in K}\sum_{i\in I}\theta_n^k r_{in}^{s}, \end{aligned} \tag{21}$$

where T^l is the total traffic obtained from local caching. T^d is the total traffic obtained from D2D links. T^b is the total traffic obtained from AP-terminal links. T^s is the total traffic obtained from the LEO satellite caching server. The total system power costs that terminals consume by MEC caching is computed as

$$\begin{aligned} &Power \\ &= P_{ca} + P_{trans} \\ &= \sum_{n\in N}\sum_{k\in K}\theta_n^k\sum_{i\in I}x_{in}^{d}\sum_{i'\in I_b^i}p_d + Zp_{ca}\sum_{n\in N}\sum_{k\in K}\sum_{i\in I}\frac{c_{kn}^{d}}{a} \\ &\quad + \sum_{n\in N}\sum_{k\in K}\theta_n^k\sum_{i\in I}\sum_{j'\in J_b^i}x_{in}^{b}y_{ij'}P_k^b + Zp_{ca}\sum_{n\in N}\sum_{k\in K}\sum_{j\in J}\frac{c_{kn}^{b}}{a} \\ &\quad + \sum_{n\in N}\sum_{k\in K}\theta_n^k\sum_{i\in I}x_{in}^{s}P_n^s + Zp_{ca}\sum_{n\in N}\frac{c_n^s}{a}, \end{aligned} \tag{22}$$

With the virtualization of network resources, different resource properties and performance metrics in the networks have different effects on the system, which are regarded as system benefits and costs, and the target system utility function can be expressed as the subtraction of system benefits and costs to measure the system. Specifically, Tra is the traffic revenue obtained by caching, and $Power$ is the power expenditure of the system with caching. The logarithmic function with the characteristics of non-decreasing and concave is a common method in data processing, and the relative relationship of the data will not be changed after taking the logarithm [59]. Therefore, logarithmic functions are used for both benefits and costs in the objective function as $\log Tra$ and $\log Power$, respectively, and the system function is shown as

$$U = \alpha\log Tra - \phi\log Power, \tag{23}$$

where α represents the revenue coefficient per unit of received traffic and ϕ represents the cost coefficient per unit of power consumption. And for the LEO satellite, C3 is the constrain of cache capacity representing the limited storage. C9 is the constrain of caches coded blocks to avoid duplicate caching. The joint optimization problem is shown as

$$\begin{aligned} &\max_{(c_{kn}^d, c_{kn}^b, c_n^s, y_{ij})} \alpha\log Tra - \phi\log Power. \\ &s.t.\ C1: \sum_{n\in N}c_{kn}^{d} \le L, \quad \forall k\in K. \\ &\quad C2: \sum_{n\in N}c_{kn}^{b} \le L^b, \quad \forall k\in K. \\ &\quad C3: \sum_{n\in N}c_{n}^{s} \le L^s. \\ &\quad C4: \sum_{k\in K}\sum_{i\in I}\sum_{j\in J}y_{ij}P_k^b \le P_{\max}^{b}, \quad \forall i\in I,\ \forall j\in J. \\ &\quad C5: y_{ij}\in[0,1], \quad \forall k\in K,\ \forall i\in I,\ \forall j\in J. \\ &\quad C6: \sum_{n\in N}\theta_n^k = 1, \quad \forall k\in K. \\ &\quad C7: c_{kn}^{d} \le a, \quad \forall k\in K,\ \forall n\in N. \\ &\quad C8: c_{kn}^{b} \le a, \quad \forall k\in K,\ \forall n\in N. \\ &\quad C9: c_{n}^{s} \le a, \quad \forall n\in N. \end{aligned} \tag{24}$$

IV. Mayfly Algorithm-Based Joint Optimization of Caching Placement and Power Allocation

Inspired by the social behavior of mayfly, especially their mating process, some researchers have proposed the mayfly algorithm [50] to solve optimization problems. The immature mayflies take several years to grow to rise to the water surface as adults. An adult mayfly only survives for a few days until it completes the purposes of reproduction. The male mayflies attract the female mayflies through a unique up and down movement pattern at a few meters above the water. When the mating is completed, the female mayflies drop their eggs on the water, and their life cycle is over. In order to use the mayfly algorithm to solve the objective optimization problem, the transformation of the original problem is necessary.

To transform constrained optimization problems into unconstrained optimization problems, we utilize the concept of

Algorithm 1 Mayfly Algorithm-Based Joint Optimization of Caching Placement and Power Allocation

1: **Input:** The size of the mayfly population S, offspring population R, convergence criteria ε, and bounds of decision variables based on constraints $C5$, $C7$, $C8$, and $C9$.
2: Initialize a set of solutions in the feasible region of the objective function, corresponding to male and female mayflies, respectively.
3: **Repeat**
4:   **for** $f \in S$ **do**
5:     Update v_f^{e+1} based on (33).
6:     Update $po\left(\hat{c}, y\right)_f^{e+1}$ based on (32).
7:     Evaluate solutions based on (26).
8:   **end for**
9:   **for** $m \in S$ **do**
10:     Update v_m^{e+1} based on (29).
11:     Update $po\left(\hat{c}, y\right)_m^{e+1}$ based on (28).
12:     Evaluate solutions based on (26).
13:     Update the local optimal solution based on (30).
14:     Update the global optimal solution based on (31).
15:   **end for**
16:   **for** $offspring \in R$ **do**
17:     Mate the mayflies based on (34) and (35), evaluate the offsprings.
18:   **end for**
19:   Create new population, keep the best males and the best females.
20: **Until** the convergence criteria condition is satisfied.
21: **Output:** The optimization results of the objective variables and the system utility function.

penalty method. The conversion process is described as follows. The penalty function is defined as

$$P\left(c,y\right)=\vartheta\sum g\left(f\left(c,y\right)\right)f^{2}\left(c,y\right) +\xi\sum g\left(h\left(c,y\right)\right)h^{2}\left(c,y\right), \quad (25)$$

where ϑ and ξ are penalty terms. (c,y) is the optimization variable group. The equality constraints are transformed into the form of $h\left(c,y\right)=0$. When $h'\left(c,y\right)=0$, the constraint is satisfied and then $g\left(h\left(c,y\right)\right)=0$, which means the new objective function value remains the same as before. $h'\left(c,y\right)\neq 0$ shows that the constraint is not satisfied and then $g\left(h\left(c,y\right)\right)=1$. The value of the new objective function is greatly increased because of the penalty term. In this way, the unexpected solutions can be removed. The inequality constraints are transformed into the form of $f\left(c,y\right)\leq 0$ with the same principle. The integer variable c is converted into a continuous variable $\hat{c}$ for subsequent optimization. It is selected according to the utility value of the two integers with the closest optimization value when restoring. Hence, the original problem (23) can be transformed into (26), shown at the bottom of the page.

Finally, the optimization problem is given as

$$\begin{aligned} &\min \phi\left(\hat{c},y\right). \\ &s.t.\ y_{ij}\in\left[0,1\right], \quad \forall k\in K, \quad \forall i\in I,\ \forall j\in J. \\ &\quad \hat{c}_{kn}^{d}\leq a, \quad \forall k\in K,\ \forall n\in N. \\ &\quad \hat{c}_{kn}^{b}\leq a, \quad \forall k\in K,\ \forall n\in N. \\ &\quad \hat{c}_{n}^{s}\leq a, \quad \forall n\in N. \end{aligned} \quad (27)$$

It is assumed that the mayfly is already an adult after hatching from the egg and the most suitable mayfly survives. The position of each mayfly represents a potential solution for the problem. First of all, there are two random sets of mayflies representing male population and female population, which means the mayflies are placed in the problem space randomly.

Denote v as the velocity of the mayfly. $po\left(\hat{c},y\right)$ is used to describe the optimization object. The position of mayflies in the problem space is equivalent to the potential solution of the problem. The gathering of male mayflies in groups means that the position of each male mayfly is adjusted according to the experience of itself and its neighbors. Let $po\left(\hat{c},y\right)_m^e$ represent the position of male mayfly m in iteration e. By adding velocity v_m^{e+1} to the current position, the movement process of male mayfly is shown as

$$po\left(\hat{c},y\right)_m^{e+1}=po\left(\hat{c},y\right)_m^{e}+v_m^{e+1}. \quad (28)$$

It is assumed that the speed of male mayflies is not very fast and they continue to move. Then the velocity of male mayfly is defined as

$$v_m^{e+1}=v_m^{e}+a_1e^{-\beta r_p^2}\left(pbest_m-po\left(\hat{c},y\right)_m^{e}\right) +a_2e^{-\beta r_g^2}\left(gbest-po\left(\hat{c},y\right)_m^{e}\right), \quad (29)$$

where v_m^e represents the velocity of mayfly m in iteration e. a_1 is the constant for scaling the cognitive contribution and a_2 is used to scale the social component contribution. $pbest_m$ represents the best position of mayfly m in history and $gbest$ represents the global best position. β is a fixed visibility factor to limit the visibility of the mayfly. r_p is the Cartesian distance between the mayfly current position and $pbest_m$. r_g is the Cartesian distance between the mayfly current position

$$\begin{aligned}\phi\left(\hat{c},y\right)&=-U+P\left(\hat{c},y\right)\\ &=-U+\xi\sum_{k\in K}g\left(\sum_{n\in N}\hat{c}_{kn}^{d}-L\right)\left(\sum_{n\in N}\hat{c}_{kn}^{d}-L\right)^2+\xi\sum_{k\in K}g\left(\sum_{n\in N}\hat{c}_{kn}^{b}-L^{b}\right)\left(\sum_{n\in N}\hat{c}_{kn}^{b}-L^{b}\right)^2\\ &\quad+\xi g\left(\sum_{n\in N}\hat{c}_{n}^{s}-L^{s}\right)\left(\sum_{n\in N}\hat{c}_{n}^{s}-L^{s}\right)^2+\xi\sum_{j\in J}g\left(\sum_{k\in K}\sum_{i\in I}y_{ij}P_k^b-P_{\max}^b\right)\left(\sum_{k\in K}\sum_{i\in I}y_{ij}P_k^b-P_{\max}^b\right)^2\end{aligned} \quad (26)$$

and $gbest$. And

$$pbest_m = \begin{cases} po(\hat{c},y)_m^{e+1}, if \phi\left(po(\hat{c},y)_m^{e+1}\right) < \phi(pbest_m). \\ is \quad kept, others. \end{cases} \tag{30}$$

$$gbest = \min\left\{\phi\left(pbest_1\right), \phi\left(pbest_2\right), \ldots, \phi\left(pbest_M\right)\right\}, \tag{31}$$

where M is the total number of male mayflies.

The movement process of female mayfly is shown as

$$po\left(\hat{c},y\right)_f^{e+1} = po\left(\hat{c},y\right)_f^e + v_f^e. \tag{32}$$

The best female is attracted by the best male, and the second best female is attracted by the second best male. The process is shown as

$$v_f^{e+1} = \begin{cases} v_f^e + a_2 e^{-\beta r_{mf}^2}\left(po\left(\hat{c},y\right)_m^e - po\left(\hat{c},y\right)_f^e\right), \\ if \phi\left(po\left(\hat{c},y\right)_f^e\right) > \phi\left(po\left(\hat{c},y\right)_m^e\right). \\ \\ v_f^e + fe * r, \\ if \phi\left(po\left(\hat{c},y\right)_f^e\right) \leq \phi\left(po\left(\hat{c},y\right)_m^e\right). \end{cases} \tag{33}$$

where r_{mf} is the distance between male mayfly and female mayfly. fe is a random walk coefficient used when females are not attracted by males. $r \in [-1,1]$ is a random value.

The method of producing offsprings also follows that the best female mates with the best male, the second best female mates with the second best male, and so on. The result of the crossover is two offsprings, which are generated as follows

$$offspring1 = off * male + (1 - off) * female, \tag{34}$$

$$offspring2 = off * female + (1 - off) * male, \tag{35}$$

where off is a random value within a range and the initial speed of offsprings is 0. Then the mayfly algorithm-based optimization of caching placement and power allocation can be described in Algorithm 1.

The process of simulating mayfly courtship dance and random flight enhances the balance between exploration and exploitation properties, which are two effective methods that can help the algorithm escape from local optima. The iteration number of the algorithm is set as $I_{\max}$, the size of the population is set as S, and d is the data dimension. The complexity of the algorithm is $O\left(I_{\max} * S * d\right)$.

V. Simulation Results and Discussions

In this section, the proposed scheme is simulated and the performance of the scheme is demonstrated with different settings. We consider an area covered by a macro base station, served by a single LEO satellite. The APs and terminals are distributed randomly in the area. The covering radius of the macro base station and APs is 1000 m and 200 m, respectively. The transmit power of APs and terminals is 20 dBm and 10 dBm, respectively [16]. The terrestrial bandwidth is 20 MHz for downlink as well as uplink. The bandwidth is 400 MHz for satellite-terrestrial links. The AWGN power spectral density is set as -174 dBm/Hz [52]. From the perspective of MVNOs, the indicators for evaluating a system are mainly determined by the income and expenditure of the system [60], [61], and the units are often unified as "unit". To keep the same magnitude of revenue and cost in the system, the corresponding coefficients are defined according to the properties and magnitude of the resources. $\alpha = 10^{12}$ units per block and $\phi = 5 \times 10^{10}$ units per Watt.

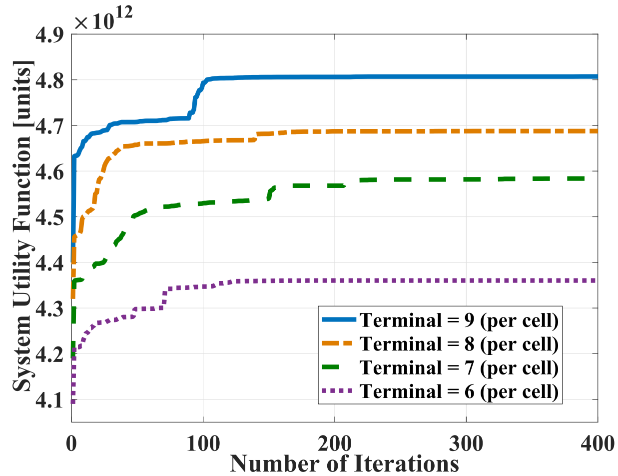


Fig. 3. System utility function versus the iterations with different system parameters.

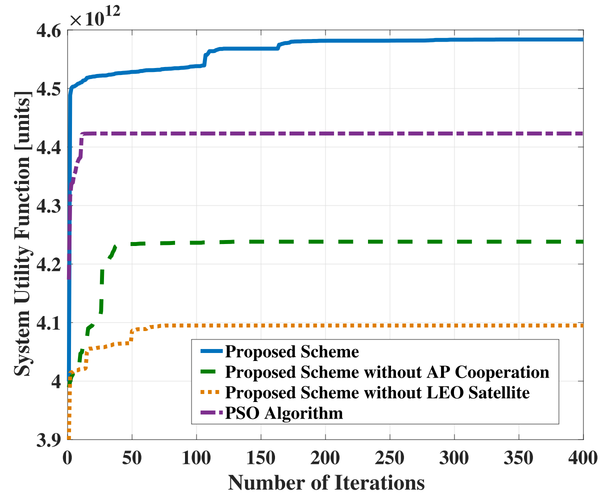
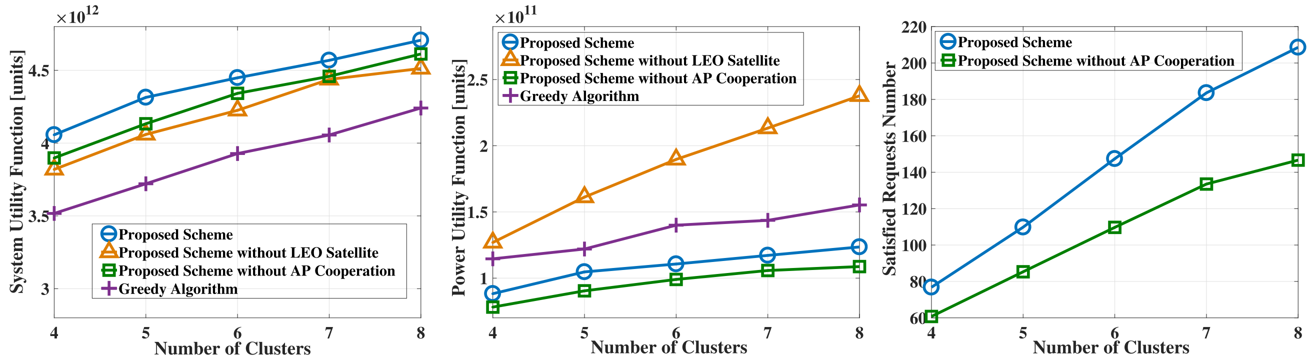


Fig. 4. System utility function versus the iterations in different schemes.

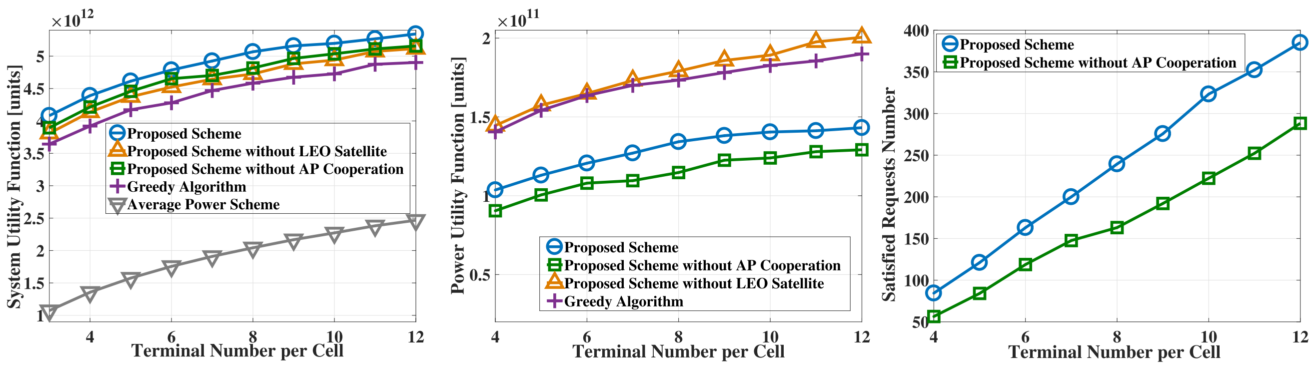
The performance is evaluated from the aspects of the number of terminals, the number of clusters, the content size of MEC servers, and the charging price of power. Meanwhile, the proposed LEO satellite-terrestrial cooperative caching scheme is compared with the scheme without the LEO satellite and the scheme without AP cooperation. The algorithm is also compared with the random and greedy algorithm for power allocation and caching placement. In addition, the performance is measured by total system utility function and power consumption utility function.

Fig. 3 shows the relationship between total system utility function and iterations based on the mayfly algorithm with different parameters. The system utility function of the system gradually stabilizes with the increasing number of iterations.



(a) System utility function vs. cluster number. (b) Power costs vs. cluster number. (c) Satisfied requests vs. cluster number.

Fig. 5. System performance versus the number of clusters in different schemes.



(a) System utility function vs. terminal number. (b) Power costs vs. terminal number. (c) Satisfied requests vs. terminal number.

Fig. 6. System performance versus the number of terminals per cell in different schemes.

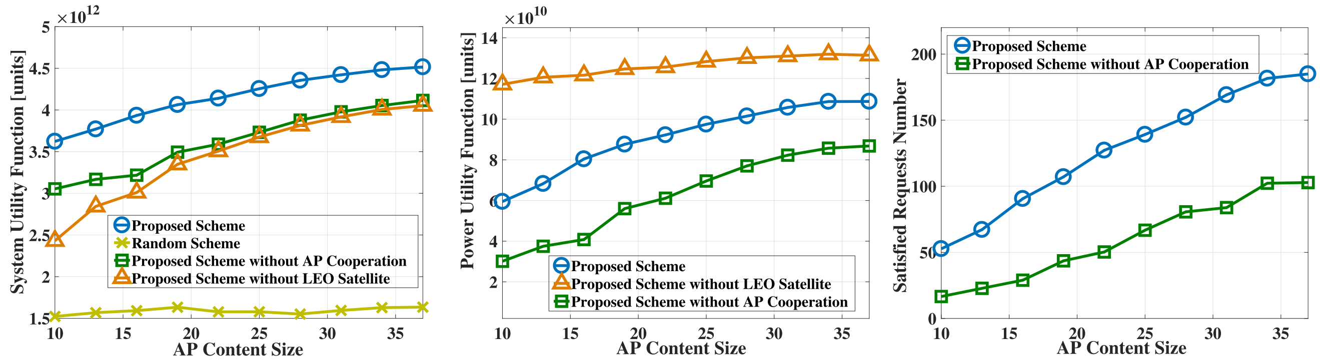
The stability value obtained with different system parameters is also different. The system utility function tends to go up with the number of terminals. These results prove the effectiveness of the algorithm for the optimization problem.

Fig. 4 shows the relationship between total system utility function and iterations in different schemes. Based on the mayfly algorithm, the proposed LEO satellite-terrestrial cooperative caching scheme has better performance. With the same system conditions, the system performance of the scheme without the LEO satellite is the worst and that of the scheme without AP cooperation is also inferior to the proposed scheme, which proves the effectiveness of the proposed satellite-terrestrial cooperative caching scheme. In addition, the particle swarm optimization (PSO) is used to optimize the proposed cooperative caching scheme under the same conditions. The mayfly algorithm used in the paper shows stronger performance in the optimization process. It is easier for the mayfly algorithm to get rid of local optima due to the process of wedding dance and random flight.

Fig. 5 (a) shows the system utility function and power costs in different schemes and algorithms with changing clusters. The total system utility function presents an increasing trend with the rising number of clusters. The proposed scheme performs better compared with others in the total utility function. Fig. 5 (b) shows the power costs with changing clusters. The power costs of the scheme are lower than the scheme without LEO satellite and the greedy algorithm, but slightly higher than the scheme without AP cooperation. In view of this situation, we analyze the number of satisfied requests between the two schemes in Fig. 5 (c), and find that there is a terrible performance of satisfying requests for the scheme without AP cooperation. With the same system conditions, less requests are satisfied without AP cooperation, which means that less traffic offloading links are established. As a result of few valid connections, the power consumption of transmitting data is saved correspondingly.

Fig. 6 (a) shows the total system utility function in different schemes and algorithms with increasing terminals per cell. With the increase of terminals, more connections are established within the system affordability and the total system utility function increases. The proposed scheme demonstrates better performance compared with others. Fig. 6 (b) shows the power costs in different schemes and algorithms with increasing terminals per cell. The reason of the slightly higher power costs compared with the scheme without AP cooperation is the same as that in Fig. 5.

Fig. 7 (a) and Fig. 7 (b) show the total system utility function and power costs in different schemes and algorithms with different content sizes of APs. The total system utility function presents an upward trend with the increasing content size. With the increasing memory of the cache servers, more coded blocks of files can be cached, which means that more requests of terminals can be met and the offloading traffic goes up. Meanwhile, the power consumed



(a) System utility function vs. AP content size.
(b) Power costs vs. AP content size.
(c) Satisfied requests vs. AP content size.

Fig. 7. System performance versus the content size of APs in different schemes.

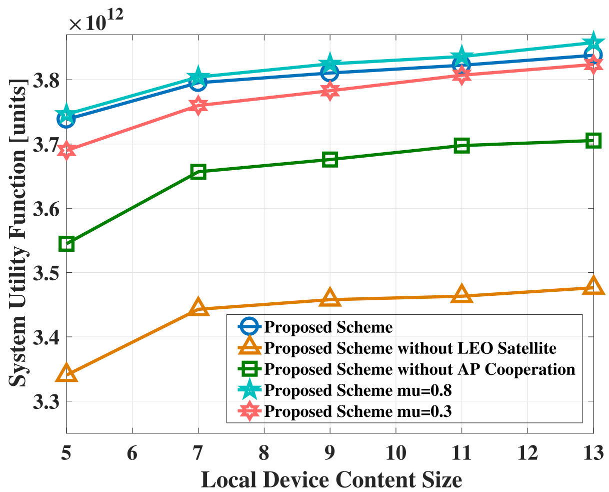


Fig. 8. System utility function versus the content size of terminals under different schemes.

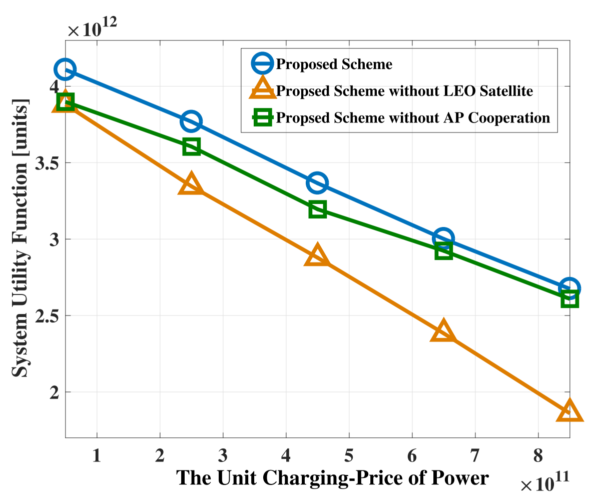


Fig. 9. System utility function versus the charging-price of power under different schemes.

by caching also increases. The proposed scheme performs well in the system utility function compared with other schemes. Fig. 7 (c) explains the reason of the nonoptimal power costs in Fig. 7 (b).

Fig. 8 shows the total system utility function with different content sizes of local devices in different schemes. Most popular files account for the majority of requests. μ is a parameter used to characterize the popularity of files in Zipf distribution. The larger the value is, the more the content is reused, which can also be seen in the simulation results. The total system utility function presents an upward trend with the increasing of the content size of local devices. The proposed scheme works better than other schemes in the system utility function.

Fig. 9 illustrates the impact of charging price of power on the total system utility function in different schemes. With the increase of the charging price of power consumption, the utility function from power costs occupies a larger proportion, which leads to the decline in overall system utility function. The proposed scheme also shows well performance.

VI. Conclusion

In this paper, we investigated the joint optimization of caching placement and power allocation in virtualized satellite-terrestrial networks. A virtualized satellite-terrestrial network framework was proposed to integrated caching resources and communication resources firstly. The optimization problem was described as the sum of revenue and expenditure of the communication system and was reformulated to facilitate the algorithm design. An intelligent optimization algorithm based on the mayfly algorithm was adopted to obtain the solution. The process of courtship movements and random flights was utilized to break away from the local optimal solution. The simulation results showed that the algorithm could converge within an amount of iterations and the proposed scheme worked better compared with other schemes in system utility function.

References

[1] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 36–43, May 2014.

[2] M. Shafi et al., "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[3] M. Hirzallah, M. Krunz, B. Kecicioglu, and B. Hamzeh, "5G new radio unlicensed: Challenges and evaluation," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 3, pp. 689–701, Sep. 2021.

[4] Z. Xiao, X. Dai, H. Jiang, and D. Wang, "Vehicular task offloading via heat-aware MEC cooperation using game-theoretic method," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 2038–2052, Mar. 2020.

[5] Z. Ding, J. Xu, O. A. Dobre, and V. Poor, "Joint power and time allocation for NOMA-MEC offloading," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6207–6211, Mar. 2019.

[6] C.-Y. Li et al., "Transparent AAA security design for low-latency MEC-integrated cellular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3231–3243, Mar. 2020.

[7] L. Shi, Y. Ye, X. Chu, and G. Lu, "Computation energy efficiency maximization for a NOMA-based WPT-MEC network," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10731–10744, Jul. 2021.

[8] K. Guo, R. Gao, W. Xia, and T. Q. S. Quek, "Online learning based computation offloading in MEC systems with communication and computation dynamics," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1147–1162, Feb. 2021.

[9] W.-K. Lai, Y.-C. Wang, H.-C. Lin, and J.-W. Li, "Efficient resource allocation and power control for LTE-A D2D communication with pure D2D model," *IEEE Trans. Veh. Technol.*, vol. 69, no. 3, pp. 3202–3216, Mar. 2020.

[10] G. Zhao, S. Chen, L. Qi, L. Zhao, and L. Hanzo, "Mobile-traffic-aware offloading for energy- and spectral-efficient large-scale D2D-enabled cellular networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3251–3264, Jun. 2019.

[11] Z. Ning et al., "Mobile edge computing enabled 5G health monitoring for Internet of Medical Things: A decentralized game theoretic approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 12, pp. 463–478, Feb. 2021.

[12] H. Zhang, Y. Qiu, X. Chu, K. Long, and V. C. M. Leung, "Fog radio access networks: Mobility management, interference mitigation, and resource optimization," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 120–127, Dec. 2017.

[13] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," 2022, *arXiv:2206.00422*.

[14] A. M. Ibrahim, A. A. Zewail, and A. Yener, "Device-to-device coded-caching with distinct cache sizes," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 2748–2762, May 2020.

[15] N. Deng and M. Haenggi, "The benefits of hybrid caching in Gauss-Poisson D2D networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 6, pp. 1217–1230, Jun. 2018.

[16] H. Zhang, H. Liu, J. Cheng, and V. C. M. Leung, "Downlink energy efficiency of power allocation and wireless backhaul bandwidth allocation in heterogeneous small cell networks," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1705–1716, Apr. 2018.

[17] R. Deng, B. Di, H. Zhang, L. Kuang, and L. Song, "Ultra-dense LEO satellite constellations: How many LEO satellites do we need?" *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 4843–4857, Aug. 2021.

[18] C. Qi, Y. Yang, R. Ding, S. Jin, and D. Liu, "Multibeam satellite communications with energy efficiency optimization," *IEEE Commun. Lett.*, vol. 26, no. 4, pp. 887–891, Apr. 2022.

[19] K. An, T. Liang, G. Zheng, X. Yan, Y. Li, and S. Chatzinotas, "Performance limits of cognitive-uplink FSS and terrestrial FS for Ka-band," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 5, pp. 2604–2611, Oct. 2019.

[20] S. Chen, S. Sun, and S. Kang, "System integration of terrestrial mobile communication and satellite communication—The trends, challenges and key technologies in B5G and 6G," *China Commun.*, vol. 17, no. 12, pp. 156–171, Dec. 2020.

[21] M. R. Bhatnagar and M. K. Arti, "Performance analysis of AF based hybrid satellite-terrestrial cooperative network over generalized fading channels," *IEEE Commun. Lett.*, vol. 17, no. 10, pp. 1912–1915, Oct. 2013.

[22] M. K. Arti, "Two-way satellite relaying with estimated channel gains," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2808–2820, Jul. 2016.

[23] C. Qi, H. Chen, Y. Deng, and A. Nallanathan, "Energy efficient multicast precoding for multiuser multibeam satellite communications," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 567–570, Apr. 2020.

[24] B. Di, H. Zhang, L. Song, Y. Li, and G. Y. Li, "Ultra-dense LEO: Integrating terrestrial-satellite networks into 5G and beyond for data offloading," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 47–62, Jan. 2019.

[25] J. Li, H. Lu, K. Xue, and Y. Zhang, "Temporal netgrid model-based dynamic routing in large-scale small satellite networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6009–6021, Jun. 2019.

[26] K. An et al., "Secure transmission in cognitive satellite terrestrial networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 3025–3037, Nov. 2016.

[27] X. Xiaoyi, W. Chunhui, and J. Zhonghe, "Design, analysis and optimization of random access inter-satellite ranging system," *J. Syst. Eng. Electron.*, vol. 31, no. 5, pp. 871–883, Oct. 2020.

[28] H. Tsuchida et al., "Efficient power control for satellite-borne batteries using Q-learning in low-earth-orbit satellite constellations," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 809–812, Jun. 2020.

[29] Z. Lin et al., "Refracting RIS-aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 3717–3724, Aug. 2022.

[30] J. Ye and Y.-J. Zhang, "Pricing-based resource allocation in virtualized cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 7096–7107, Jul. 2019.

[31] D. A. Ravi, V. K. Shah, C. Li, Y. T. Hou, and J. H. Reed, "RAN slicing in multi-MVNO environment under dynamic channel conditions," *IEEE Internet Things J.*, vol. 9, no. 6, pp. 4748–4757, Mar. 2022.

[32] Y. Li et al., "Understanding the ecosystem and addressing the fundamental concerns of commercial MVNO," *IEEE/ACM Trans. Netw.*, vol. 28, no. 3, pp. 1364–1377, Jun. 2020.

[33] C. Qi and X. Wang, "Precoding design for energy efficiency of multibeam satellite communications," *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1826–1829, Sep. 2018.

[34] Z. Lin, M. Lin, J.-B. Wang, T. de Cola, and J. Wang, "Joint beamforming and power allocation for satellite-terrestrial integrated networks with non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 3, pp. 657–670, Jun. 2019.

[35] Z. Jia, M. Sheng, J. Li, D. Zhou, and Z. Han, "VNF-based service provision in software defined LEO satellite networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6139–6153, Sep. 2021.

[36] J. Du, C. Jiang, J. Wang, Y. Ren, S. Yu, and Z. Han, "Resource allocation in space multiaccess systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 2, pp. 598–618, Apr. 2017.

[37] Q. Tang, Z. Fei, B. Li, and Z. Han, "Computation offloading in LEO satellite networks with hybrid cloud and edge computing," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 9164–9176, Jun. 2021.

[38] D. Zhou, M. Sheng, Y. Wang, J. Li, and Z. Han, "Machine learning-based resource allocation in satellite networks supporting Internet of Remote Things," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6606–6621, Oct. 2021.

[39] O. Kodheli, N. Maturo, S. Chatzinotas, S. Andrenacci, and F. Zimmer, "NB-IoT via LEO satellites: An efficient resource allocation strategy for uplink data transmission," *IEEE Internet Things J.*, vol. 9, no. 7, pp. 5094–5107, Apr. 2022.

[40] T. Ma et al., "UAV-LEO integrated backbone: A ubiquitous data collection approach for B5G Internet of Remote Things networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 11, pp. 3491–3505, Nov. 2021.

[41] A. Ivanov, R. Bychkvo, and E. Tcatcorin, "Spatial resource management in LEO satellite," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15623–15632, Dec. 2020.

[42] C. Han, A. Liu, L. Huo, H. Wang, and X. Liang, "A prediction-based resource matching scheme for rentable LEO satellite communication network," *IEEE Commun. Lett.*, vol. 24, no. 2, pp. 414–417, Feb. 2020.

[43] J. Chen, H. Wu, P. Yang, F. Lyu, and X. Shen, "Cooperative edge caching with location-based and popular contents for vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 10291–10305, Sep. 2020.

[44] F. Guo, H. Zhang, H. Ji, X. Li, and V. C. M. Leung, "An efficient computation offloading management scheme in the densely deployed small cell networks with mobile edge computing," *IEEE/ACM Trans. Netw.*, vol. 26, no. 6, pp. 2651–2664, Dec. 2018.

[45] Z. Zhao et al., "A novel framework of three-hierarchical offloading optimization for MEC in industrial IoT networks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5424–5434, Aug. 2020.

[46] A. Mseddi, W. Jaafar, H. Elbiaze, and W. Ajib, "Joint container placement and task provisioning in dynamic fog computing," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10028–10040, Dec. 2019.

[47] S. G. Domanal, R. M. R. Guddeti, and R. Buyya, "A hybrid bio-inspired algorithm for scheduling and resource management in cloud environment," *IEEE Trans. Services Comput.*, vol. 13, no. 1, pp. 3–15, Jan. 2020.

[48] R. Andreotti, T. Wang, V. Lottici, F. Giannetti, and L. Vandendorpe, "Resource allocation via max–min goodput optimization for BIC-OFDMA systems," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2412–2426, Jun. 2016.

[49] K. M. Sim and W. H. Sun, "Ant colony optimization for routing and load-balancing: Survey and new directions," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 33, no. 5, pp. 560–572, Sep. 2003.

[50] K. Zervoudakis and S. Tsafarakis, "A mayfly optimization algorithm," *Comput. Ind. Eng.*, vol. 145, Jul. 2020, Art. no. 106559.

[51] A. J. Gonzalez, G. Nencioni, A. Kamisiński, B. E. Helvik, and P. E. Heegaard, "Dependability of the NFV orchestrator: State of the art and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3307–3329, 4th Quart., 2018.

[52] J. Li, K. Xue, D. S. L. Wei, J. Liu, and Y. Zhang, "Energy efficiency and traffic offloading optimization in integrated satellite/terrestrial radio access networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2367–2381, Apr. 2020.

[53] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722–3732, Sep. 2016.

[54] Z. Yang et al., "Cache placement in two-tier HetNets with limited storage capacity: Cache or buffer?" *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5415–5429, Nov. 2018.

[55] J. I. Kokkala, D. S. Krotov, and R. J. Patric, "On the classification of MDS codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6485–6492, Dec. 2015.

[56] V. Bioglio, F. Gabry, and I. Land, "Optimizing MDS codes for caching at the edge," in *Proc. IEEE GLOBECOM*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

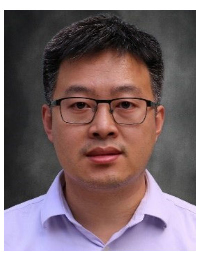
[57] X. Xu and M. Tao, "Modeling, analysis, and optimization of caching in multi-antenna small-cell networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5454–5469, Nov. 2019.

[58] Y. Xu and S. Mao, "A survey of mobile cloud computing for rich media applications," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 46–53, Jun. 2013.

[59] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[60] Z. Tan, F. R. Yu, X. Li, H. Ji, and V. C. M. Leung, "Virtual resource allocation for heterogeneous services in full duplex-enabled SCNs with mobile edge computing and caching," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1794–1808, Feb. 2018.

[61] G. O. Boateng, D. Ayepah-Mensah, D. M. Doe, A. Mohammed, G. Sun, and G. Liu, "Blockchain-enabled resource trading and deep reinforcement learning-based autonomous RAN slicing in 5G," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 1, pp. 216–227, Mar. 2022.



Haijun Zhang (Fellow, IEEE) was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia (UBC), Canada. He is currently a Full Professor and an Associate Dean with the School of Computer and Communications Engineering, University of Science and Technology Beijing, China. He is a Distinguished Lecturer of IEEE. He received the IEEE ComSoc Young Author Best Paper Award in 2017, the IEEE CSIM Technical Committee Best Journal Paper Award in 2018, and the IEEE ComSoc Asia-Pacific Best Young Researcher Award in 2019. He serves/served as the Track Co-Chair for VTC Fall 2022 and WCNC 2020/2021; the Symposium Chair for IEEE GLOBECOM'19; the TPC Co-Chair for IEEE INFOCOM 2018 Workshop on Integrating Edge Computing, Caching, and Offloading in Next Generation Networks; and the General Co-Chair for GameNets'16. He serves/served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING.

Jiali Xu received the B.S. degree from the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China, in 2020, where she is currently pursuing the M.S. degree. Her research interests include satellite-terrestrial networks, edge caching, and resource allocation in wireless communication.

Xiangnan Liu received the B.S. degree from the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, China, in 2019, where he is currently pursuing the Ph.D. degree. His research interests include access control, beamforming, and resource allocation in 6G wireless communication.

Keping Long (Senior Member, IEEE) received the M.S. and Ph.D. degrees from the University of Electronic Science and Technology of China, Chengdu, in 1995 and 1998, respectively. From September 1998 to August 2000, he was a Post-Doctoral Research Fellow with the National Laboratory of Switching Technology and Telecommunication Networks, Beijing University of Posts and Telecommunications (BUPT), China. From September 2000 to June 2001, he was an Associate Professor with BUPT. From July 2001 to November 2002, he was a Research Fellow with the ARC Special Research Centre for Ultra Broadband Information Networks (CUBIN), University of Melbourne, Australia. He is currently a Professor and the Dean of the School of Computer and Communication Engineering, University of Science and Technology Beijing. He has published more than 200 articles, 20 keynote speeches, and invited talks at international and local conferences. His research interests include optical internet technology, new generation network technology, wireless information networks, value added services, and secure technology of networks. He was a TPC or ISC Member of COIN 2003/04/05/06/07/08/09/10, IEEE IWCN2010, ICON2004/06, and APOC2004/06/08. He is a member of the editorial committees of sciences in China Series F and China Communications. He was awarded by the National Science Fund for Distinguished Young Scholars of China in 2007 and selected as the Chang Jiang Scholars Program Professor of China in 2008. He was the Co-Chair of the organization Committee for IWCMC2006, the TPC Chair of COIN 2005/08, and the TPC Co-Chair of COIN 2008/10.

Victor C. M. Leung (Life Fellow, IEEE) is currently a Distinguished Professor of computer science and software engineering with Shenzhen University, Shenzhen, China. He is also an Emeritus Professor of electrical and computer engineering and the Director of the Laboratory for Wireless Networks and Mobile Systems, The University of British Columbia (UBC), Vancouver, Canada. His research interests include broad areas of wireless networks and mobile systems. He has published widely in archival journals and refereed conference proceedings in these areas and several of his papers have won best paper awards. He is a fellow of the Royal Society of Canada, Canadian Academy of Engineering, and Engineering Institute of Canada. He was a recipient of the 1977 APEBC Gold Medal, Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarships from 1977 to 1981, the 2012 UBC Killam Research Prize, the IEEE Vancouver Section Centennial Award, the 2017 Canadian Award for Telecommunications Research, and the 2018 IEEE TCGCC Distinguished Technical Achievement Recognition Award. He has coauthored articles that won the 2017 IEEE ComSoc Fred W. Ellersick Prize, the 2017 IEEE SYSTEMS JOURNAL Best Paper Award, the 2018 IEEE CSIM Best Journal Paper Award, and the 2019 IEEE TCGCC Best Journal Paper Award. He is named in the current Clarivate Analytics list of "Highly Cited Researchers." He has provided leadership to the technical program committees and organizing committees of numerous international conferences. He is serving on the Editorial Boards for the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, the IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE NETWORK, and IEEE ACCESS. He has previously served on the editorial boards of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS WIRELESS COMMUNICATIONS Series and Series on *Green Communications and Networking*, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, the IEEE TRANSACTIONS ON COMPUTERS, and the IEEE WIRELESS COMMUNICATIONS LETTERS.